CPSC 340:
Machine Learning and Data Mining

Conclusion

Admin

- **Assignments**:
  - No more of these. Congratulations!
- **Final exam**:
  - April 14 (3:30-6pm, location SUB 2201).
  - Covers Assignments 1-6, Lectures 2-31.
  - Past exams posted on GitHub.
  - Closed-book, cheat sheet: 1-page double-sided (same as midterm).
- **Office hours**:
  - Currently 14 office hours next week
  - Check the calendar for updates

Big ideas of the course

- **Supervised vs. Unsupervised** learning
  - Supervised: regression, classification; focus on prediction
  - Unsupervised: find "structure" or "patterns" in the data; clustering; dimensionality reduction
- **Fundamental tradeoff of ML**
  - under/overfitting, (cross-)validation
- **Golden rule**
  - test data should not (significantly?) influence training
- **No free lunch theorem**
  - there is no "best" ML model (so we learn lots)
- **The different decisions you need to make and their effects:**
  1. Collecting/preprocessing your data
  2. Choosing a model
  3. Choosing a loss
  4. Choosing an optimizer

Individual topics covered

- Part 1: EDA, decision trees, NB, KNN, ensembles & random forests

- Part 2: k-means, DBSCAN, hierarchical clustering, outlier detection

- Part 3: linear regression, basis & other non-linear regression, regularization (L2, L1, L0), gradient descent & SGD, logistic regression, SVM & kernels, maximum likelihood & MAP

- Part 4: PCA & variants, NMF, recommender systems, nonlinear dimensionality reduction (MDS, ISOMAP, t-SNE)

- Part 5: Neural networks, CNNs, deep learning software.

- (blue indicates that this topic appeared on one of the assignments)

Some things we didn't cover

- We did not say much about…
  - Data collection/preparation
  - Causality
  - Sequences, time series data
  - Risks & societal/ethical implications of ML
  - Large scale problems and/or distributed computing
  - Communicating your results
  - Much more…

- Despite all this we covered a lot of ground.
  - You all should feel a sense of accomplishment!

CPSC 340 vs. CPSC 540

- Goals of CPSC 340: implementing practical machine learning methods.
  - Present most of the fundamental ideas, sometimes in simplified ways.
  - Choose models that are widely-used in practice.
  - Focus on the model fitting part of the pipeline
- Goals of CPSC 540: research-level machine learning.
  - Covers complicated details/topics/models that we avoided.
  - Targeted at people with algorithms/math/stats/numerical background.
  - Goal is to be able to understand ICML/NIPS papers at the end of course.
- Example 540 topics:
  - How many iterations of gradient descent do we need?
  - What if y_i is a sentence or an image or a protein? (Graphical models and RNNs.)
  - What if data isn't IID?

Other ML-Related Courses

- CPSC 532R:
  - Probabilistic graphical models.
- CPSC 532L:
  - Deep learning for vision, sound, and language.
- STAT 406:
  - Similar/complementary topics, focus on mathematical details and applications.
- STAT 460/461:
  - Advanced statistical issues (what happens when 'n' goes to ∞?)
- STAT 5xx
  - These all cover related topics.
- EECE 592:
  - Deep learning and reinforcement learning.
- EOSC 510:
  - Similar/complementary topics, emphasis on EOSC applications.
- EOSC 550:
  - Optimization methods for deep learning.
- LIBR 559d:
  - Language and social media data.

Evaluations & surveys (10-15 minutes)

- Science course evaluation
  - https://eval.ctlt.ubc.ca/science
  - Deadline is Monday (April 9) at 11:59pm
- TA evaluations (paper)
  - Mainly for the tutorial TAs: Clement Fung, Angad Kalra
  - Other TAs: Tanner Johnson, Michael Przystupa
  - Need 2 volunteers to take them to the CS main office (ICCS 201) after class
    - Please put all blank forms back in the envelope as well
  - Please share any pencils you have

Prizes

- I'd like to thank a few students who participated a lot in class.
  - This makes it a lot more meaningful for me.
  - I didn't announce this at the beginning because I didn't want people to participate "for the prize".
- Top Piazza answerers:
  - Winners: Oliver (most endorsed answers), Jonathan (most asked)
  - Honourable mention to Eric C, Trevin, Kyle
- Most lively in-class participants:
  - Winners: Fed, Connor
  - Honourable mention to Edwin C, Mohamed E, Tom
- Best excuse for missing class:
  - Aly, Arie, Kevin S
- Unofficial course troublemaker:
  - Fred

Grad School Advice / FAQs

- Don't do it!
  - Or, if you insist…
- Get as much work experience as possible: co-op full-time work after undergrad, grad school internships
  - Build skills, confidence, awareness in what's out there and, sometimes, *boredom* (or curiosity)
- Research Master's vs. Professional Master's
  - Research Master's ("pays you"): reading/writing papers, inventing new things, leads to PhD
  - Professional Master's (you pay): practical skills to get you a job, leads to industry
- Keep in mind that Canada is a world leader in ML
  - you don't necessarily need to go far to find a world-class research group
- Improving your graduate school applications:
  - Do interesting side-projects and post them on your GitHub profile (also great for industry)
  - Make a personal website and/or blog (also great for industry)
  - For a research Master's, prior ML research experience is a huge boost to your application
  - Pick your referees carefully
    - Knows you well in an academic/professional context >> is famous
    - Try to gauge if the person is enthusiastic about writing the letter
  - If there's something that needs explaining (e.g., low grades), explain it!
- Next steps at UBC:
  - Take CPSC 540 (open to undergrads) and other courses mentioned earlier
  - Get work experience, do projects

(Unsolicited) General Life Advice

- Try to find an intersection of work you enjoy and careers with enough jobs
  - It's up to you to determine both of those things
  - Do your own research, make your own decisions
  - Don't let your parents influence you too much
- Don't obsess over grades (gamification)
  - You should know when/why you need good grades; they are not worth anything inherently
  - They do not reflect the skills needed for success (not even close)
  - Don't measure peoples' worth by grades: having a C+ vs. A+ GPA doesn't make you "inferior" or "superior"
- Make sure you're happy in the present moment
  - Don't sacrifice happiness because you're "working towards something"
  - You should enjoy university!
- Don't assume the system makes sense
  - For example, undergraduate achievement has little to do with graduate school or research achievement
  - Your education may not focus on the skills you need to succeed in 2017-2100 (we try, but it's hard)
  - Older or more "senior" people (like me) **can be wrong**!
- You are lucky
  - UBC is one of the best schools in Canada
  - But the above applies even in the best places ☺

Thank you

- Thank you for being a great audience!